**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
**C12N 15/12, C12Q 1/68, G01N 33/53**

**A1**

(11) International Publication Number: **WO 99/05274**

(43) International Publication Date: 4 February 1999 (04.02.99)

(21) International Application Number: PCT/US98/15272

(22) International Filing Date: 23 July 1998 (23.07.98)

(30) Priority Data:
08/898,838 23 July 1997 (23.07.97) US

(71) Applicants *(for all designated States except US)*: SMITHKLINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US). SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Great West Road, Brentford, Middlesex TW8 9EP (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BERGSMA, Dirk, Jon [US/US]; 271 Irish Road, Berwyn, PA 19312 (US). WILSON, Shelagh [GB/GB]; Beckets Bramfield, Hertford, Hertfordshire SG14 QQJ (GB).

(74) Agents: HAN, William, T. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406–0939 (US).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

**Published**
*With international search report.*

(54) Title: METHOD OF IDENTIFYING NOVEL G–PROTEIN RECEPTORS AND THEIR FUNCTIONS

(57) Abstract

The present application relates to methods of identifying new 7–transmembrane (G–protein coupled) receptors and methods for identifying ligands and receptors' biological functions.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Faso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

# METHOD OF IDENTIFYING NOVEL G-PROTEIN RECEPTORS AND THEIR FUNCTIONS

## FIELD OF INVENTION

The present application relates to methods of identifying new 7-transmembrane (G-protein coupled) receptors and methods for identifying ligands and receptors' biological functions.

## BACKGROUND OF INVENTION

The advent of rapid DNA sequencing spawned the 'genomic era', leading to the initiation of the Human Genome Project. The novel technologies developed in association with genomic research have already had a significant impact on the way investigations into the basis of disease are being conducted and will, no doubt, substantially enhance how diseases are diagnosed and treated in the near future. To keep pace with the evolution of molecular medicine, the pharmaceutical industry has embraced genomics and is attempting to exploit the new technologies to identify novel targets for drug discovery. The major questions which remain to be addressed concern how to convert genomic sequences into therapeutic targets in an expeditious manner and eventually obtain pharmaceuticals to enhance the quality of life. This invention relates to G protein-coupled receptors (GPCRs), particularly to so-called 'orphan' receptors[1].

G protein-coupled receptors are a superfamily of integral plasma membrane proteins involved in a broad array of signaling pathways. Since the first cloning of GPCR gene sequences over a decade ago, novel members of the GPCR superfamily continue to emerge through cloning activites as well as bioinformatic analyses of sequense databases, although their ligands are unidentified and their physiological/biological function (relevance) remain to be defined. These 'orphan' receptors provide a rich source of potential drug discovery targets.

The GPCR superfamily is related both structurally and functionally. The signature motif of these receptors is seven distinct hydrophobic domains, which are 20 to 30 amino acids in length, that are linked by hydrophilic amino acid sequences of varied lengths[2,3]. Biophysical[4] and biochemical[5] studies support the notion that these receptors are intercalated into the plasma membrane with the N-terminus extracellular and the C-terminus in the cytoplasmic portion of the cell. Therefore, these receptors are often referred to as seven

transmembrane receptors or 7TM receptors. While it is not yet known how many individual genes actually encode these receptors, it is clear that this family of proteins is one of the largest yet identified. Functionally, GPCRs share in common the property that upon agonist binding they transmit signals across the plasma membrane through an interaction with heterotrimeric G proteins[6,7]. These receptors respond to a vast range of agents[2,5,8] such as protein hormones, chemokines, peptides, small biogenic amines, lipid-derived messengers, divalent cations (a calcium sensor has been identified that is a GPCR[9]) and even proteases such as thrombin, which activates its receptor by cleaving off a portion of the N-terminus[10]. Finally, these receptors play an important role in sensory perception including vision and smell[2,5,8]. Correlated with the broad range of agents that activate these receptors is their existence in a wide variety of cells and tissue types and thus they play roles in a diverse range of physiologic processes. It is likely, therefore, that the GPCR superfamily is involved in a variety of pathologies. This point was recently emphasized by the surprizing discovery that certain GPCRs for chemokines act as cofactors for HIV infection[11-13].

GPCRs represent the primary mechanism by which cells sense alterations in their external environment and convey that information to the cells' interior. Binding of an agonist to the receptor promotes conformational changes in the cytoplasmic domains leading to the interaction of the receptor with its cognizant G protein(s). Agonist-promoted coupling between receptors and G proteins leads to the activation of intracellular effectors which substantially amplify second messenger production feeding into the signaling cascade. Since effectors are often enzymes (e.g. adenylyl cyclase[14] which converts ATP to cyclic AMP or phospholipase C[15] which hydrolyzes inositol lipids in membranes to release inositol trisphosphate, which in turn mobilizes calcium within a cell) or ion channels[16], many second messenger molecules can be produced as the result of a single agonist binding event with its receptor. Changes in the intracellular levels of ions and/or cyclic AMP result in modulation of distinct phosphorylation cascades[17,18], extending through the cytosol to the nucleus, that eventually culminate in the physiological response of the cell to the extracellular stimulus. Although the overall paradigm is apparently the same for all GPCRs, the diversity of receptors, G proteins and effectors suggest a myriad of potential signaling processes and this becomes an important concept for identification of the function of orphan GPCRs.

To date more than 800 GPCRs have actually been cloned from a variety of eukaryotic species, from fungus to humans[19]. For humans, the most represented species, about 140 GPRCs have been cloned for which the cognate ligands are also known. This number excludes the sensory olfactory receptors, which are predicted to number in hundreds to thousands. By traditional molecular genetic approaches, coupled with the explosion in genomic information, we have been able to identify more than 100 additional orphan GPCR family members. By definition, there is enough sequence information in the receptor cDNAs to clearly place them in the superfamily of G protein-coupled receptors, but there is often not sufficient sequence homology with known members of this family to be able to assign their ligands with confidence and/or predict their function. In total, there are currently over 240 human GPCRs, excluding sensory receptors. As the size of sequence databases continue to increase this list could grow to 400, and perhaps even to 1000 or more unique gene products. The list will grow even further as paralogs and alternatively spliced variants emerge. Most orphan GPCRs share a low degree of sequence homology, typically about 25-35% overall amino acid sequence identity, with any known GPCRs, suggesting that they belong to new subgroups of receptors. Indeed, several orphans show closer homology to each other than to known GPCRs. Nevertheless, the majority of orphan receptors are phylogenetically distributed among a broad spectrum of distantly-related known receptors subgroups.

GPCRs have a proven history of being excellent therapeutic targets. Within the last 20 years several hundred new drugs have been registered which are directed towards activating or antagonizing GPCRs and it is estimated that a majority of current research within the pharmaceutical industry is focussed on this signaling pathway[20]. Table 1 shows a representative snapshot of a variety of receptors, disease targets, and corresponding drugs. It is clear from this table that the therapeutic targets span a wide range of disorders and disease states.

TABLE I

| GPCR | Generic | Drug | Indication |
|---|---|---|---|
| Acetylcholine Muscarinic | Bethanechol | Urecholine | GI |
| | Dicyclomine | Bentyl | GI |
| | Ipratropium | Atrovent | CP |
| | | | |
| Adrenoceptor | | | |
| b1 | Atenolol | Tenormin | CP |
| a2 | Clonidine | Catapres | CP |
| b1/b2 | Propranolol | Inderal | CP |
| a1 | Terazosin | Hytrin | CP |
| b2 | Albuterol | Ventolin | CP |
| b1/b2/a1 | Carvedilol | Coreg | CP |
| | | | |
| AngiotensinII | Losartan | Cozaar | CP |
| | Eprosartan | Teveten | CP |
| | | | |
| Calcitonin | Calcitonin | Calcimar | Osteoporosis |
| | eel-Calcitonin | Elcatonin | Osteoporosis |
| | | | |
| Dopamine | | | |
| D2 | Metoclopramide | Reglan | GI |
| D2/D3 | Ropinirole | Requip | CNS |
| D2 | Haloperidol | Haldol | CNS |
| | | | |
| Histamine | | | |
| H1 | Dimenhydrinate | Dramamine | CNS |
| H1 | Terfenadine | Seldane | CP |
| H2 | Cimetidine | Tagamet | GI |
| H2 | Ranitidine | Zantac | GI |

| | | | |
|---|---|---|---|
| Gonadatropin Releasing Factor | Goserelin | Zoladex | Cancer |
| | Nafarelin | Synarel | Endometriosis |
| | | | |
| Leukotriene | Pranlukast | Onon | CP |
| | Zafirlukast | Accolate | CP |
| | | | |
| Opioid | | | |
| Kappa | Buprenorphine | Buprenex | CNS |
| | Butorphanol | Stadol | CNS |
| Mu | Alfentanil | Alfenta | CNS |
| | Morphine | Kadian | CNS |
| | | | |
| Oxytocin | | Syntocinon | Labor |
| | | | |
| Prostaglandin | Epoprostenol | Flolan | CP |
| | Misoprostol | Cytotec | GI |
| | | | |
| Somatostatin | Octreotide | Sandostatin | Cancer |
| | | | |
| Serotonin | | | |
| 5-HT1D | Sumatriptan | Imitrex | CNS |
| 5-HT2A | Ritanserin | Tisertan | CNS |
| 5-HT4 | Cisapride | Propulsid | GI |
| 5-HT1B | Trazodone | Desyrel | CNS |
| 5-HT2A/2C | Clozapine | Clozaril | CNS |
| | | | |
| Vasopressin | Desmopressin | | CP/Renal |

Abbreviations: CNS, central nervous system; CP, cardiopulmonary; GI,

Another example of the significance and versatility of GPCRs is the number of cases of genetic diseases that are linked to defects in these proteins. It is likely that many more genetic diseases will be mapped to GPCRs receptors as the era of genomics continues to expand and families with inherited mutations are examined much more comprehensively. Clearly there is a need for identification and characterization of further 7TM receptors which can play a role in preventing, ameliorating or correcting dysfunctions or diseases.

The importance of GPCRs to drug discovery continues to be manifested by the fact that across the pharmaceutical industry active research projects, ranging from basic studies all the way through to advanced development, are focused on GPCRs as primary targets. Molecular biology has had a dramatic influence on these efforts. The cloning of cDNAs for well known GPCRs led to the discovery of a surprising number of paralogs[5]. These novel receptor subtypes were unexpected because the current cornucopia of pharmacological agents did not possess the required selectivity to clearly distinguish all of them and thus an oppotunity for drug discovery was quickly recognized. Current research efforts seek to define the physiology associated with these novel receptor subtypes and to discover highly selective compounds as potential pharmaceuticals. These efforts are almost exclusively focused on GPCRs for which activating ligands are known. Since characterized GPCRs were, and continue to be attractive therapeutic targets, it is most reasonable to speculate that many of the orphan receptors will have a similar potential. The method of the present invention involves determination of the function of these orphan receptors, through the identification of activating ligands and, once the function is clarified, link the orphan receptors to a specific disease and thus establish it as a candidate for a full fledged drug discovery effort.

As used herein, an orphan GPCR refers to a novel G-protein coupled receptor for which there is not sufficient sequence homology (identity) with any other known members of G-protein coupled receptor family to be able to assign its ligand(s) and thus unable to predict its function. The two receptors are said to have low sequence homology (identity) or not sufficient sequence homology (identity), when homology (identity) is under 60%, preferably under 50%, but more preferably between 35-20%.

"Identity" is a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using

published techniques. See, e.g.: (COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, 1988; BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, 1993; COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, 1987; and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipton, D., *SIAM J Applied Math* (1988) 48:1073. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCS program package (Devereux, J., *et al.*, *Nucleic Acids Research* (1984) 12(1):387), BLASTP, BLASTN, FASTA (Atschul, S.F. *et al.*, *J Molec Biol* (1990) 215:403).

## SUMMARY OF THE INVENTION

This invention relates to a method for identifying drug discovery targets or function of orphan GPCRs, which comprises:

(i) analyzing the structures of a pool of partial or full length gene sequences to identify those partial or full length genes that encode putative GPCRs based on 7-transmembrane receptor motifs, preferably based on bioinformatics;

(ii) expressing the full length genes in recombinant host cells suitable for ligand fishing;

(iii) screening for the ligands (natural or surrogate) by ligand fishing; and

(iv) inferring the function of the putative GPCR based on the characteristics of ligands that bind to it and thereby identifying those putative GPCRs that are useful as drug discovery targets.

In one embodiment the pool of partial or full length gene sequences employed in (i) includes a number of gene sequences that are contigs assembled from partial gene sequences.

In another embodiment the full length genes expressed in step (ii) were first identified as partial genes encoding putative GPCRs and then fully cloned prior to expression in recombinant host cells.

Yet in another embodiment the partial or full length genes are pre-selected based on the types of tissues in which the genes are expressed and/or on chromosome mapping prior to expression in the recombinant host cells.

Yet in further embodiment the ligand fishing step, step (iii), includes multiple different functional or binding assays.

Further embodiment involves ligand fishing be carried out against known GPCR ligands, against extracts from tissues, biological fluids and cells, and/or against compounds, including synthetic peptides, in a compound or combinatorial library.

Yet further embodiment involves that the potential ligands for the ligand fishing step are selected from the same tissue type as those from which the putative GPCRs were derived.

Yet in further embodiment involves that the function of the putative receptor is inferred in step (iv) from binding to a known GPCR ligand or by first determining the biological effects of the ligand if the ligand is not a known GPCR ligand.

The invention also provides a method of using the newly discovered ligands by the methods above to generate antibodies thereto and thereby allowing to determine further function of the receptors.

## BRIEF DESCRIPTION OF FIGURES

Figure 1. Paradigm shift from classical to reverse molecular phamacological approaches to drug discovery

Figure 2. Strategy for utilizing orphan GPCRs as targets for drug discovery

## DETAILED DESCRIPTION

### "Reverse" Molecular Pharmacology

Until recently, research on the identification of GPCRs as targets for drug discovery has been conducted with a traditional approach as illustrated in Figure 1A. For this strategy, one usually starts with a functional activity, which forms the basis of an assay by which a ligand is identified through purification from biological fluids, cell supernatants, or tissue extracts. One example of the success of this strategy is the discovery of the potent vasoconstricting peptide, endothelin[21]. Once isolated, the ligand is used to characterize its cellular and tissue biology as well as its pathophysiological role. Subsequently, cDNAs encoding corresponding receptors are 'fished' from gene libraries using a variety of methodologies (e.g., receptor purification and expression cloning) that often either directly or indirectly use the ligand as the 'hook'. As the nucleotide sequences for GPCRs began to accumulate and be analysed, additional receptors were cloned by homology screening, by polymerase chain reaction (PCR) methodologies which employed oligonucleotide primers based on nucleotide sequences conserved within the seven transmembrane domains of the GPCR family and positional cloning. Once the cloned human receptor cDNA is expressed in a heterologous cell system[22], it together with its ligand are used to form the basis of a screen to explore chemical compound libraries for receptor antagoninsts or agoninsts. Lead structures identified in the screen are refined through medicinal chemistry using an iterative process. Resulting drug leads with appropriate *in vivo* pharmacology are passed on into the clinic for development. In summary, this traditional scheme starts with a functional activity which provides an assay for the purification of a ligand and subsequent identification of its receptor. The receptor and cognate ligand are then used in a screen that ultimately aids in the discovery and design of a novel drug.

Recently, the applicants were able to devise a radically different departure from the traditional approach with the introduction of a new "reverse" molecular pharmacological strategy, shown diagramatically in Figure 1B.

Through both traditional molecular cloning techniques and, more recently, mass sequencing of expressed sequence tags (ESTs) from cDNA libraries, it is now possible to identify G protein-coupled receptors through computational, or bioinformatic methodologies. The EST approach, initially proposed by Sidney Brenner (University of Cambridge) and first brought to large scale practice by Craig Venter (The Institute of Genome Research),

constitutes random, single pass sequencing of cDNAs randomly picked from a collection of cDNA libraries, followed by extensive bioinformatic analysis of the sequence to identify structural signatures characteristic of GPCRs. Once new members of the GPCR superfamily are identified, the receptors themselves are used as the 'hook' in functional assays (e.g. using calcium, cAMP, microphysiometer, oocyte electrophysiology, etc.) to fish for natural ligands in tissue extracts of human, and other mammalian, species, such as porcine tissue. Specifically such tissue extracts include lung, liver, gut, heart, kidney, adrenals, ischemic brain, plasma, urine and placenta. Extracts that produce positive functional responses can be sequencially subfractionated until an activating ligand is isolated and identified. The receptor/ligand pair are then used for compound bank screening to identify a lead compound that, together with the activating ligand, is used for biology/ pathophysiology studies to determine function and the potential therapeutic value of a receptor antagonist (or agonist) to ameliorate a disease process. Further evaluation of therapeutic potential can involve chromosomal mapping studies of the receptor, together with identification of receptor-associated genetic markers which will allow genotyping of disease populations. Once a disease link is finally identified, an appropriate compound can be advanced for clinical study.

The concept of the above "reverse" molecular pharmacological strategy was validated by the success fully decribed in applicants' copending applications [USSN 08/846,705 filed April 30, 1997; USSN 08/846,704 filed April 30, 1997; and 08/887,382 filed July 2, 1997, which is a continuation-in-part application of USSN 08/820,519 filed March 19, 1997, which further claims priorty to USSN 60/033,604 filed December 17, 1996. All these applications are incorporated herein by reference in their entirety]. Briefly, a partial orphan receptor HFGAN72 sequence was initially identified through EST analysis method involving a computer database search, and subseqeuntly full length cloning was achieved. The receptor was expressed in HEK 293 cells and ligand fishing was successfully performed to discover natural ligands from bovine hypothalamus and rat brain tissue extracts. Results from *in situ* hybridization on adult rat brain slices showed that HFGAN 72 receptor ligands are strongly expressed in both the hypothalamus and in the hypothalamal neurons. Since the location of HFGAN 72 receptor ligands are localized in hypothalamus, it was immediately inferred that they have a number of neurological and psychiatric implications. Subsequent rat feeding study confirmed that HFGAN 72 receptor ligand may be an endogenous regulator of appetite, and that antagonists of its receptor may be useful in the treatment of obsesity and diabetes,

whilst agonists or antagonists may be useful in the treatment of eating disorders such as anorexia nervosa, bulimia, and cachexia, among others.

**Screening Strategy**:

Figure 2 illustrates the generic strategy for "reverse" molecular pharmacological approach. In addition to the EST approach, which has yielded the majority of our collection of orphan receptors, we have also utilized a number of more traditional approaches such as low stringency screening, using portions of known GPCRs as hybridization probes, as well as PCR-based methods. By these techniques we have succeeded in identifying more than 70 orphan receptors in addition to those already in the public domain.

Since cDNAs identified by EST cloning are often incomplete, Northern hybridization analysis is used to establish the tissue or cell pattern of mRNA expression of the GPCRs. This information is used to identify the tissue/cell cDNA libraries which are to be probed for full length clones, and significantly, to determine whether a receptor is expressed in a particular disease target tissue of interest. A highly selective tissue expression pattern may also provide a clue as to receptor function. Once obtained, full length GPCR clones are expressed in mammalian cell lines and yeast model systems (see below) for functional analysis. *Xenopus* oocytes may also be used for expression; however low screening throughput limits their use to a secondary, confirmatory assay system. For mammalian cell expression, the HEK 293 cell line or CHO cells are frequently used. These cell types possess a large repertoire of G proteins which would be necessary for coupling to downstream effectors. They also share a solid history of positive functional coupling for a wide variety of known GPCRs. However, since receptor coupling can not be accurately predicted from primary sequence data, orphan GPCRs may need to be expressed in a variety of cell lines to establish viable coupling.

These heterologous expression systems form the basis for screening for an activating ligand. The success of establishing functional coupling of the recombinant receptor depends to a large extent on whether the receptor is properly expressed, which may be assessed by Northern or Western blot analysis, and whether appropriate G proteins and downstream effectors are present in the cell in which the receptor is expressed. Before ligand fishing is initated, several steps need to be taken. Because it is difficult to accurately predict the coupling specificity of orphan GPCRs from their primary sequence, assays must be chosen

which will detect a wide range of coupling mechanisms. These generally focus on changes in intracellular levels of cAMP or $Ca^{2+}$, but can also include more generic measurements such as metabolic activation of the cell via the cytosensor microphysiometer[23]. Recently, it has become possible to configure most of these screens in high throughput format by employing fluorescent-based assays and using charge-coupled device cameras and reporter gene constructs that allow easy readout in microtiter plate format. Ever increasing throughput of the assays will be nesssary to screen large libraries as demanded by competitive drug discovery. However, this approach is somewhat cumbersome and inefficient if all the assays described above have to be employed. It may be possible to funnel heterologous signal transduction through a defined pathway The prospect of a single transduction pathway assay was raised by the observation that heterologous expression of the G protein subunit, Ga15/16, promoted coupling of various GPCR subfamily members through activation of phospholipase Cb and likely calcium mobilization24,25. The diversity of the GPCRs sucessfully coupled through $G_{a16}$ to phospholipid metabolism suggests this is a useful method to screen for orphan receptor activation.

Once heterologous receptor expression is achieved and functional assays are in place, ligand fishing experiments can be initiated. Although the homology with known GPCRs is low, one can, if so desired, begin by screening the orphans against known GPCR ligands; since the sequence homology between some subtypes of known receptors can be low (e.g. 30-40% between neuropeptide Y receptor subtypes), it is possible that new paralog receptors for known ligands still remain to be discovered. The next step is to search for novel activating ligands by screening biological extracts obtained from tissues, biological fluids and cell supernatants. An additional option is screening libraries of compounds for activating ligands. Complex libraries of peptides or compound collections could be rich sources of 'surrogate' agonists which would promote receptor activation and coupling but are not endogenous ligands. Rationale for searching for surrogate agonists springs from a report that a nonpeptide agonist has been discovered for the angiotensin II receptor[26]. There is slao an obvious precedent for nonpeptide agonists for opioid receptors. These data suggest that surrogate agonists need not mimic endogenous agonist binding exactly to initiate signaling events. Screening of the very large libraries which will be generated by fractionation of biological extracts and by combinatorial chemical synthesis require that the functional assays

used be not only high throughput but also robust as false positives can be a significant problem.

Examples are beginning to emerge which show progress has been made in characterizing novel GPCRs. A first example is the identification of a G protein-coupled receptor which functions as a calcitonin gene-related peptide (CGRP) receptor[27]. CGRP is a 37 amino acid peptide, widely distributed in neurons, and functions as a potent vasodilator. It may be involved in migraine and has been implicated in Type II diabetes by promoting insulin resistance.

A novel GPCR EST was derived from a human synovium cDNA library[27]. Sequence analysis showed the new GPCR to have about 56% similarity to the human calcitonin receptor and was hence originally expected to be a new subtype of the calcitonin receptor. The message for this novel receptor was expressed predominantly in lung, which is known to be a relatively rich source of CGRP receptors. Following full length cloning from a human lung library, the receptor cDNA was stably expressed in HEK293 cells. Both radioligand binding using $^{125}$I-CGRP, as well as functional assays of CGRP-stimulated cAMP accumulation, demonstrated an appropriate pharacological profile for the expressed receptor similar to that observed with endogenous CGRP receptors on human neuroblastoma cells. Using a similar set of approach, other novel receptors such as C3a receptor[28] and CCK5 receptor were identified.

Further, two groups recently investigated an opioid-like receptor, ORL-1[29,30]. Both groups expressed the GPCR in Chinese hamster ovary (CHO) cells and challenged the transfected cells with a series of opiate agonists, but without response. Both groups then employed a similar ligand fishing approach. Taking crude extracts from rat brain[29] or porcine brain[30], they screened against the stably transfected cell lines using inhibition of adenylyl cyclase actitivity as a functional assay. They were able to fractionate the brain extracts and identify the novel dynorphin-like ligand, which they called nociceptin[29] or orphaninFQ[30]. Thus, both teams successfully established a functional assay in transfected CHO cells that allowed the purification of a novel 17 amino acid neuropeptide ligand for the novel receptor.

Importantly, however, all the examples given above are for receptors with significant homology to other known GPCR superfamily members, and their activating ligands proved to

be known GPCR ligands or ligands from tissue extracts suspected of containing such ligands because the novel receptor had significant of homology to other known GPCRs for which biological function can be inferred. This invention relates to a novel method for identifying drug discovery targets or function of orphan GPCR, which comprises:

(i) analyzing the structures of a pool of partial or full length gene sequences to identify those partial or full length genes that encode putative GPCRs based on 7-transmembrane receptor motifs, preferably based on bioinformatics;

(ii) expressing the full length genes in recombinant host cells suitable for ligand fishing;

(iii) screening for the ligands (natural or surrogate) by ligand fishing; and

(iv) inferring the function of the putative GPCR based on the characteristics of ligands that bind to it and thereby identifying those putative GPCRs that are useful as drug discovery targets.

REFERENCES cited above:


1 Libert, F., Vassart, G. and Parmentier, M. (1991) *Current Opinion in Cell Biology* **3**, 218-223

2 Strader, C. D., Fong, T. M., Tota, M. R. and Underwood, D. (1994) *Annual Reviews Biochem.* **63**, 101-32

3 Baldwin, J. M. (1994) *Current Opinion in Cell Biology* **6**, 180-190

4 Schertler GFX, Villa C., Henderson, R. (1993) *Nature* **362**, 770-772

5 Dohlman, H. G., Thorner, J., Caron, M. G. and Lefkowitz, R. J. (1991) *Annual Reviews Biochem.* **60**, 653-688

6 Neer, E. J., (1995) *Cell,* **80**, 249-257

7 Rens-Domiano, S. and Hamm, H. E. (1995) *FASEB J.* **9**, 1059-1066

8 Coughlin, S. R., (1994) *Current Opinion in Cell Biology* **6**,191-197

9 Brown, E. M., Gamba, G., Riccardi, D., Lombardi, M., Butters, R., Kiford, O, Sun, A., Hediger, M. A., Lytton, J. and Herbert, S. C.(1993) *Nature* **366**, *575-580*

10 Vu, T-K H., Hung, D. T. Wheaton, V. I., Coughlin, S. R.(1991) *Cell* **64**, 1057-1068

11 Feng, Y., Broder, C. C., Kennedy, P. E. and Berger, E. A. (1996) *Science* **272**, 872-876

12 Deng, H. K., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., DiMarzio, P., Marmon, S. Sutton, R. E., Hill, C. M., Davis, C. B., Peiper, S. C. Schall, T. J., Littman, D. R. and Landau, N. R. (1996) *Nature* **381**, 661-666

13 Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P. and Paxton, W. A. (1996) *Nature* **381**, 667-673

14 Sunahara, R. K., Dessauer, C. W. and Gilman, A. G. (1996) *Annu. Rev. Pharmacol. Toxicol.* **36**, 461-480

15 Rhee, S. G. and Choi, K. D. (1992) *Advances in Second Messenger and Phosphoprotein Research* **26**, 35-49

16 Clapham, D. E. (1995) *Cell* **80**, 259-268

17 Hunter, T. (1995) *Cell* **80**, 225-236

18 Graves, J. D., Campbell, J. S. and Krebs, E. G. (1995) *Annals of the NY Academy of Sciences* **766**, 320-341

19 Kolakowski, L. F. (1997) **GCRDb-WWW The G protein-Coupled Receptor DataBase World-Wide-Web Site** http://receptor.mgh.harvard.edu/GCRDBHOME.html.org

20 Roush, W. (1996) *Science* **271**, 1056-1058

21 Yanagisawa, M., Kurihara, H., Kimura, S., Tomobe, Y., Kobayashi, M., Mitsui, Y., Yazaki, Y., Goto, K., and Masaki, T. (1988) *Nature* **332**, 441

22 Tate, C. G. and Grisshammer, R. (1996 *Trends in Biotechnology* **14**, 426-430

23 McConnell, H.M., Owicki, J.C., Parce, J.W., Miller, D.L., Baxter, G.T., Wada, H.G. and Pitchford, S. (1992) *Science* **257**, 1906-1912

24 Offermanns, S. and Simon, M., (1995) *Journal of Biological Chemistry* **270**, 15175-15180

25 Milligan, G., Marshall, F. and Rees, S. (1996) *Trends in Pharmacological Sciences* **17**, 235-237

26 Perlman, S., Schambye, H. T., Rivero, R. A. Greenlee, W. J., Hjorth, S. A. and Schwartz, T. W. (1995) *Journal of Biological Chemistry* **270**, 1493-1496

27 Aiyar, N. Rand, K., Elshourbagy, N. A., Zeng, Zhizhen, Adamou, J. E., Bergsma, D. J. and Li, Y. (1996) *Journal of Biological Chemistry* **271**, 11325-11329

28 Ames, R. S., Li, Y., Sarau, H. M., Nuthulaganti, P., Foley, J. J., Ellis, C., Zeng, Z., Su, K., Jurewicz, A., Hertzberg, R. P., Bergsma, D. J. and Kumar, C. (1996) *Journal of Biological Chemistry* **271**, 20231-20234

29 Meunier, J-C., Mollereau, C., Toll, L., Suaudeau, C., Moisand, C., Alvinerie, P., Butour, J-L., Guillemot, J-C., Ferrara, P., Monsarrat, B., Mazargull, H., Vassart, G., Parmentier, M. and Costentin, J. (1995) *Nature* **377**, 532-535.

30 Reinscheid, R. K., Nothacker, H-P., Bourson, A., Ardati, A., Henningsen, R. A., Bunzow, J. R., Grandy, D. K., Langen, H., Monsma, F. J., Jr. and Civelli, O. (1995) *Science* **270**, 792-794.

31 Broach, J. R. and Thorner, J. (1996) *Nature* **384 Supp**, 14-16

32 Price, L. A., Kajkowski, E. M., Hadcock, J. R., Ozenberger, B. A. and Pausch, M. H. (1995) *Molecular and Cellular Biology* **15**, 6188-6195

33 Manfredi, J. P., Klein, C., Herrero, J. J., Byrd, D. R. Trueheart, J., Wiesler, W. T., Fowlkes, D. M. and Broach, J. R. (1996) *Molecular and Cellular Biology* **16**, 4700-4709

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

5

What is claimed is:

1. A method for identifying drug discovery targets or function of orphan GPCR which comprises:

(i) analyzing the structures of a pool of partial or full length gene sequences to identify those partial or full length genes that encode putative GPCRs based on 7-transmembrane receptor motifs, preferably based on bioinformatics;

(ii) expressing the full length genes in recombinant host cells suitable for ligand fishing;

(iii) screening for the ligands (natural or surrogate) by ligand fishing; and

(iv) inferring the function of the putative GPCR based on the characteristics of ligands that bind to it and thereby identifying those putative GPCRs that are useful as drug discovery targets.

2. The method of claim 1 wherein the pool of partial or full length gene sequences employed in (i) includes a number of gene sequences that are contigs assembled from partial gene sequences.

3. The method of claim 1 wherein the full length genes expressed in step (ii) were first identified as partial genes encoding putative GPCRs and then fully cloned prior to expression in recombinant host cells.

4. The method of claim 1 wherein the partial or full length genes are pre-selected based on the types of tissues in which the genes are expressed and/or on chromosome mapping prior to expression in the recombinant host cells.

5. The method of claim 1 wherein the ligand fishing step, step (iii), includes multiple different functional assays.

6. The method of claim 1 wherein ligand fishing is carried out against known GPCR ligands, against extracts from tissues, biological fluids and cells, and/or against compounds in a compound or combinatorial library

7. The method of claim 1 wherein the potential ligands for the ligand fishing step are selected from the same tissue type as those from which the putative GPCRs were derived.

8. The method of claim 1 wherein the function of the putative receptor is inferred in step (iv) from binding to a known GPCR ligand or by first determining the biological effects of the ligand if the ligand is not a known GPCR ligand.
the ligand if the ligand is not a known GPCR ligand.

9. A method of using the ligands discovered by the method of claim 1 to generate antibodies thereto and thereby allowing to determine further function of the receptors.

REFERENCES TO FIGURES 1 AND 2

NOT TO BE TAKEN INTO ACCOUNT FOR THE PURPOSE OF INTERNATIONAL PROCESSING

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/15272

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :C12N 15/12; C12Q 1/68; G01N 33/53
US CL :435/6, 7.1, 7.2; 436/501

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.1, 7.2; 436/501

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN/MEDLINE
search terms: G protein coupled, receptor#, orphan#, combinatorial, random peptide#, ligand#, immobil?, identif?.

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | O'DOWD ET AL. A novel gene codes for a putative G protein-coupled receptor with an abundant expression in brain. FEBS Letters. October 1996. Vol. 394, pages 325-329, see entire document. | 1-9 |
| Y | LACKMANN ET AL. Purification of a ligand for the EPH-like receptor HEK using a biosensor-based affinity detection approach. Proceedings of the National Academy of Science. March 1996. Vol. 93. pages 2523-2527, see entire document. | 1-9 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

| | | | |
|---|---|---|---|
| * | Special categories of cited documents: | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "A" | document defining the general state of the art which is not considered to be of particular relevance | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "E" | earlier document published on or after the international filing date | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L" | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "&" | document member of the same patent family |
| "O" | document referring to an oral disclosure, use, exhibition or other means | | |
| "P" | document published prior to the international filing date but later than the priority date claimed | | |

| | |
|---|---|
| Date of the actual completion of the international search<br>08 OCTOBER 1998 | Date of mailing of the international search report<br>23 OCT 1998 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>JOHN D. ULM<br>Telephone No. (703) 308-0196 |

Form PCT/ISA/210 (second sheet)(July 1992)*

THIS PAGE BLANK (USPTO)